

Reducing White Noise

Towards a common Information garbage policy

Dr. Stefan Gradmann

**Regionales Rechenzentrum der Universität
Hamburg**

White Noise ?

- A first look at the problem (not altogether serious):
The Infinite Monkey Protocol Suite (IMPS)
- Slightly more serious:
The prime action of use is selection, and here we are halting indeed. There may be millions of fine thoughts, and the account of the experience on which they are based, all encased within stone walls of acceptable architectural form; but if the scholar can get at only one a week by diligent search, his syntheses are not likely to keep up with the current scene..
(Vannevar Bush, As We May Think)
-

White Noise ?

-
- **A definition:**
„A sound or signal consisting of all audible frequencies with equal intensity. The term is used analogously to the term 'white light' in optics which denotes the simultaneous presence of colours of all frequencies. „
- **A poetical implementation of White Noise**
- **Some reading: Don DeLillo, White Noise (1985)**
- **And another installation**

White Noise

-
-



Information Accumulation

- More reading: David Shenk, Data Smog (1997)
- Data smog, white noise, information overload
- Data vs. Information
 - Information is about **selecting/aggregating** and not merely **accumulating** data
 - Data is much easier to **accumulate** than information to **aggregate**
 - Logic of WWW based information services has been **quantity driven** for a long time (reacting to the former sparseness of information)
- Even with **quality awareness** starting to grow there is no **quality standard** for WWW based information services
- An example

Quality! Quality?

SuperSearch III - Microsoft Internet Explorer

Adresse <http://www.robtex.com/search/query.htm>

Virtual Library **GO!**
SETUP INFO HOME

Lycos

WEB SITES [POPULAR | WEB SITES | SHOPPING]

1,407,776 Web sites were found in a search of the complete Lycos Web catalog

- Computers > Internet > WWW > Searching the Web > Directories > WWW **Virtual Library**
- Reference > Libraries > **Library** and Information Science > User Services > Reference Services > **Virtual Reference Desk**

BARNES & NOBLE

Keyword Virtual Librai GO

GO Can I go bungee jumping in Zimbabwe? Get the scoop on travel.

Search results

64,118 matches [Next 10 >](#) | [Hide summaries](#) | [Sort by date](#) | [Ungroup results](#)

GO/Infoseek

1. **Web Developer Virtual Library**
Resources for developers creating Web sites.

Yahoo! Site Matches (1 - 19 of 336)

Government > U.S. Government > [Web Directories](#)

- [U.S. Government Information Sources - WWW Virtual Library](#)

Regional > Regions > Asia > Computers and Internet > Internet > World Wide Web > Searching the Web > [Search Engines and Directories](#)

- [Asian Studies WWW Virtual Library](#) - Web directory of scholarly documents, resources and information systems concerned with Asia.

Computers and Internet > Internet > World Wide Web > Searching the Web >

alta vista:

Home

1,630,740 pages found. **Web Pages** Products Directories Images Video MP3/A

Search for: Virtual Library any language

☐ Search within these results Example: +marketing -sales +business

Related Searches:

- [alabama virtual library](#)
- [ask eric virtual library](#)
- [usf](#)
- [itt vi](#)

See reviewed sites in:

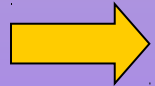
- [C Programming Resources](#)
- [Foreign Affairs & Diplomacy](#)

*** SuperSearch III ***

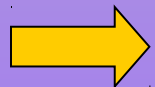
Internet

Apples and Pears

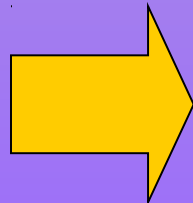
- Yahoo: 336
- InfoSeek: 64.118
- Lycos: 1.407.776
- AltaVista: 1.630.740
- Combined top 10: 35 sites (1x3, 4x2)



Great at **accumulation**, poor at **aggregation**



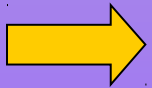
All providers operate differently: comparing the results equals comparing apples and pears



- The means to reduce white noise
- The players to bring together
- Examples of action taken/planned

•

Low Level noise reduction

- Player: **computing center**
 - Means
 - Institution-wide policies for e-mail use
 - Spam reduction
 - Proxies
 - Firewalls
 - Providing **networking services**, the infrastructure and logistics required both for noise creation and noise reduction
-  Virtual Library Unit of Hamburg University is based in the computing center

Semantic noise reduction I

White noise reduction is primarily focussed on **semantics**: information filtering and aggregation rely on notions of **relevance** and **semantic context**

- Player: **Libraries**
 - Traditionally concerned with **content selection, focussing** and **content aggregation** (content acquisition, indexing, classification, cataloguing)
 - Anchored in the traditional background of **printed material**
 - Still mostly apply information technology in a manner merely **electrifying traditional workflows**
 - Could render valuable services in reducing information overload provided they become able to **apply IT more appropriately**

Metadata Generation

- **Cataloguing vs. Metadata** (<http://www.ifla.org/IV/ifla64/007-126e.htm>)
- **Cataloguing approach**
 - Expensive
 - Time consuming
 - Inappropriate and impossible to sustain with WWW resources
- **Metadata based approach**
 - Excellent antidote against white noise (enhancing semantic precision)
 - Syntactic standardization and standard tools are lacking
 - Search engine support is lacking (which metadata are trustworthy?)
 -

Metadata: actions

- **Use metadata:** from 2001 on, top level WWW-pages in our domain (uni-hamburg.de) will contain metadata complying to DC standards (more to follow)
- **Certify:** metadata will be produced (and certified) under our university's responsibility, thus eventually minimizing the problem of trust
- **Involve staff:** metadata will be supplied by producers of scientific content – cataloguers will not have to interfere (and even though do not become obsolete)
- **Supply tools:** appropriate tools will be supplied by the computer center for all participating university staff

Semantic Information Ecology

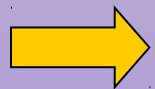
- DC-related activities result in **,syntactic' metadata** only
- They affect only resources we produce ourselves and not content we ingest from outside

 Invest in automated semantic aggregation

- **Detection, extraction** and on the fly **conversion** of metadata present in external resources
- Automated generation of **,syntactic' metadata**
- Automated generation of **,semantic' metadata** using linguistic techniques for information filtering and automated summarisation/lexical aggregation
- Building (among others) on DC-DOT, CARMEN, OAI ...

Media competence

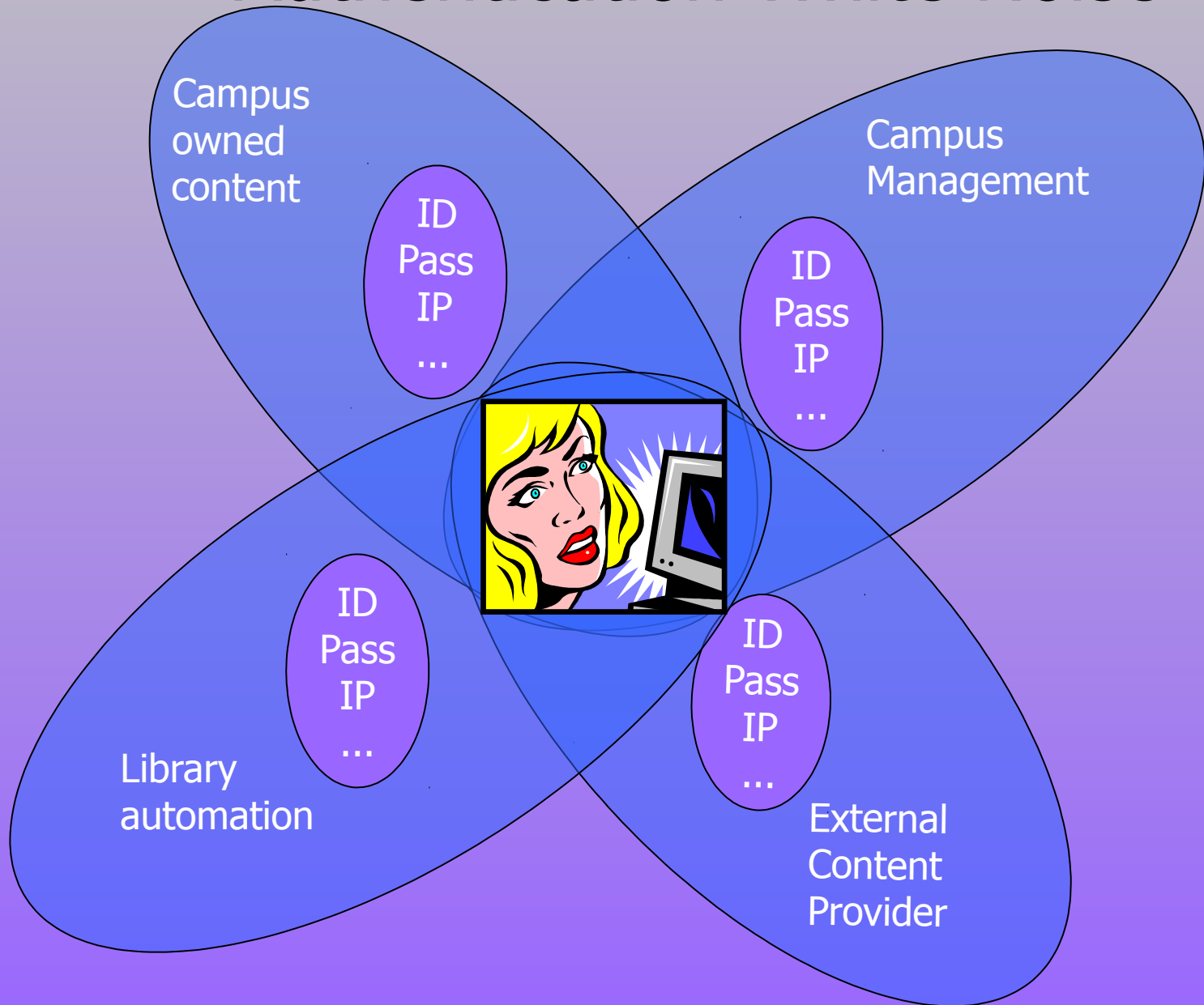
- White noise reduction via **input filtering**
- White noise reduction via **output filtering**



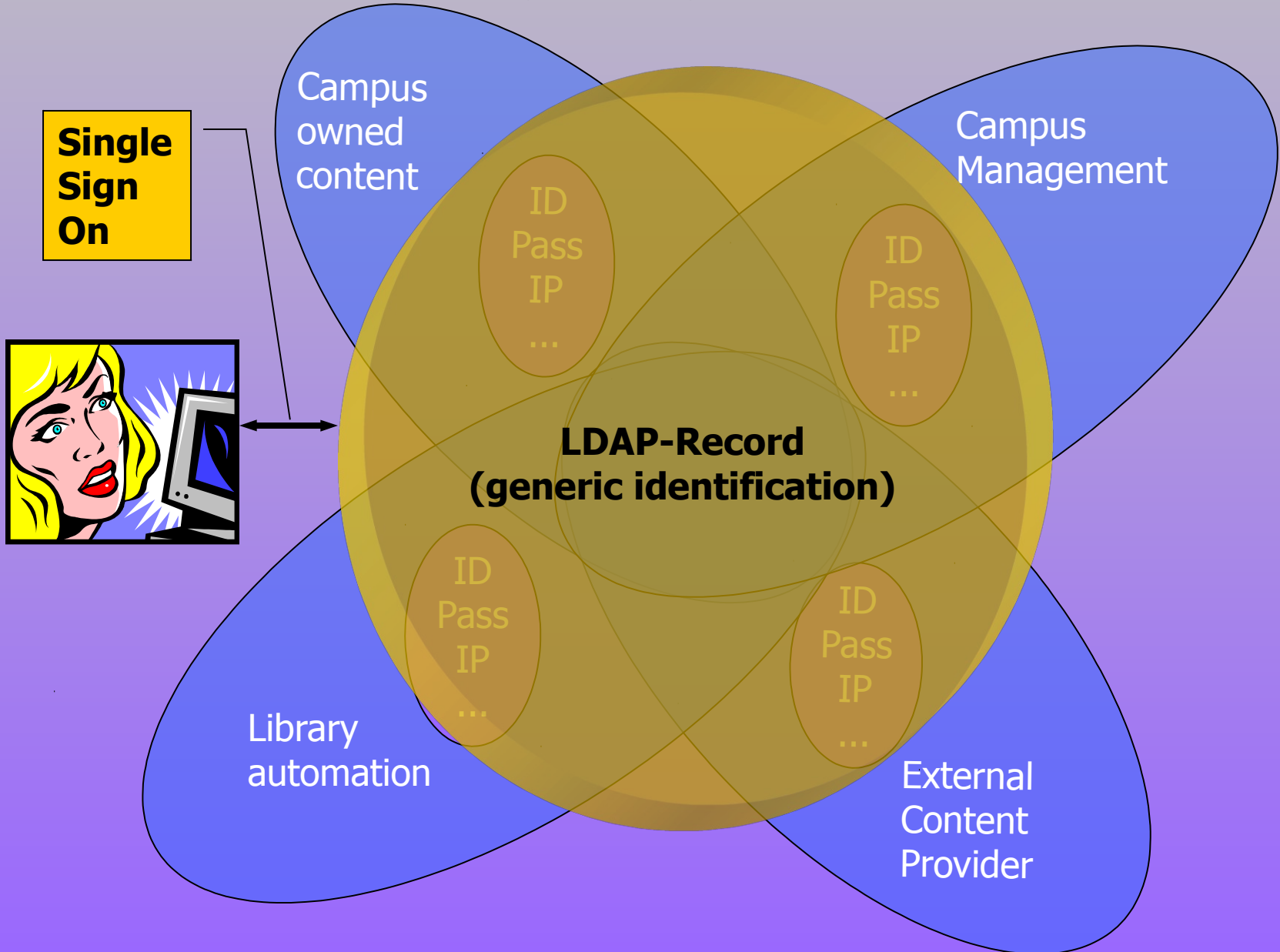
Build **media competence**

- **Center for Media Competence** is more than just a multi-media center: stress on educational aspects
 - Teach users to **make efficient use** of multimedia resources in tele-teaching and tele-learning
 - Teach users to **make sustainable use** of multi-media resources
 - Teach users to produce output in line with concepts of **information ecology**¹
- Prevent multi-media use from becoming yet another powerful means for information pollution

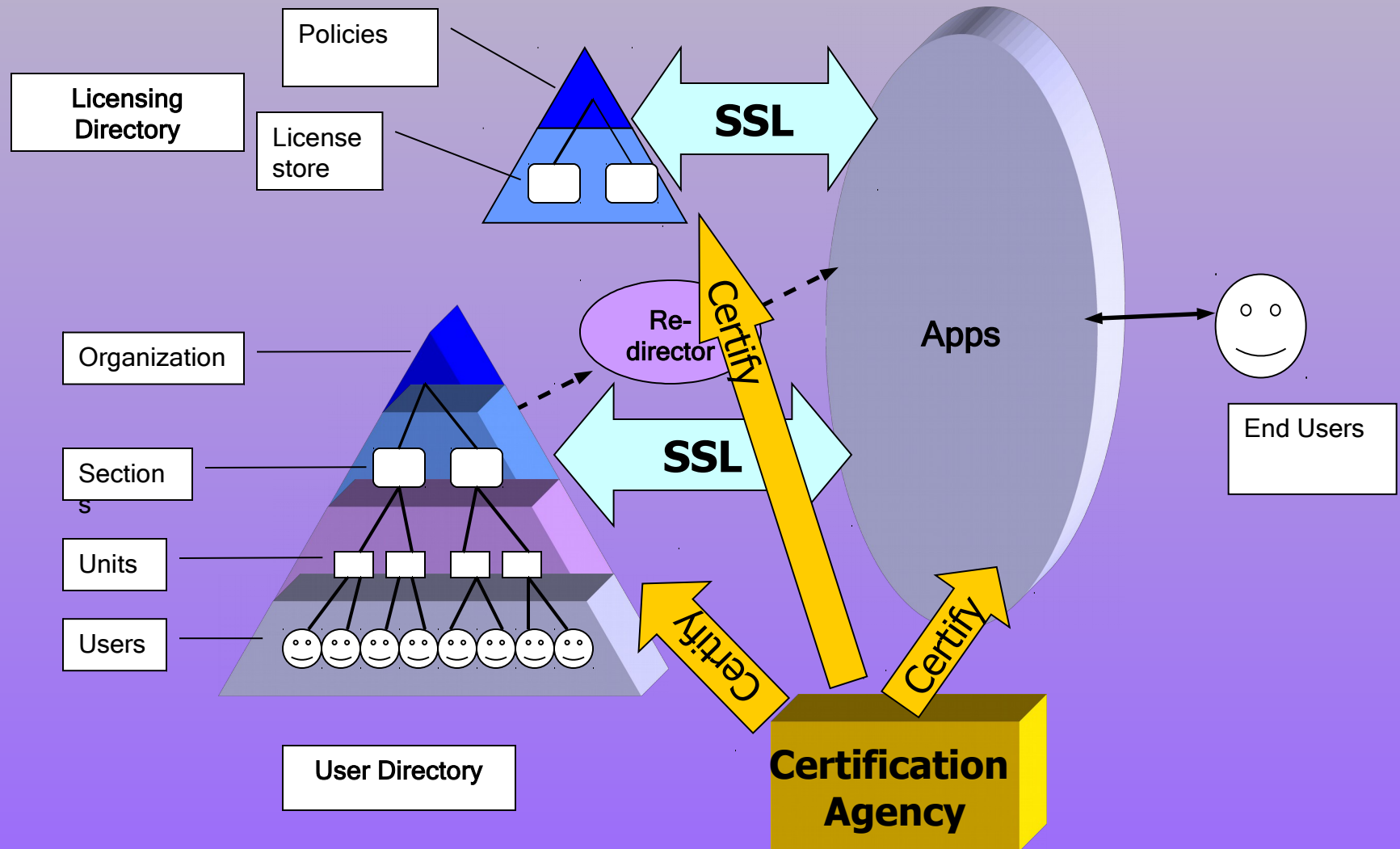
Authentication White Noise



Aggregating Authentication



Authentication Architecture



Structural Convergence II

- **University e-press**
 - Collaborative peer-reviewing
 - Collaborative metadata generation
- **WWW quality content identification (CORC)**
- **... more ...**

Reducing White Noise

- ➔ Joint efforts for with **leveraging, filtering and aggregation** of information
- ➔ As part of an institutional policy for **information aggregation** and **white noise reduction**
- ➔ Eventually resulting in **new institutional models** transcending libraries, computer and multimedia centers